INTEGRATED WEB CATALOGUE SERVICE
(IWCATS)

Serge Kéna-Cohen, Edric Keighan, Guy Rochon

Intelec Geomatics, 696, Ste-Croix, Office 200, Montreal Canada, H4L 3Y2
kena@intelec.ca

Working Group IV/4

KEY WORDS: Metadata, Spatial Warehouse, OGC

ABSTRACT:

The gathering of metadata is often a tedious and frustrating exercise that has to be performed on top of data management in spatial warehouses. Metadata standards are complex and cannot be readily applied to data managed in spatial warehouses where a data set can be defined as any object. The context for an Integrated Web Catalogue service (IWCat) is based on the postulate that in geomatics, metadata are meant to help users discover and assess geospatial information and to help providers publish their data or services.

This paper presents the results of the Integrated Web Catalogue Service (IWCatS) project. The project objectives were to develop tools that allow the integration of metadata servers in a Web environment using protocols similar to OpenGIS standards. The project results will essentially allow suppliers to devote more time to the improvement of their data thanks to the economies of time obtained in the management of their metadata. These improvements could take the form of gathering of new data, better precision or more frequent updates. The users are then able to better discover the data of interest, those will be of better quality and they will be better documented, which will ensure a more efficient use in their specific fields (forestry, mines, etc.). The IWCat service was tested with six governmental organisations involved in the management of geospatial data.

1. INTRODUCTION

Stakeholders involved in distributing and accessing geospatial data agree that the first stage in achieving this goal is the documentation of data sets or the entry of metadata. Organisations handling standards (FGDC, ISO/TC211, OpenGIS) have identified metadata as a priority in their work and geospatial data infrastructures such as the Canadian Geospatial Data Infrastructure (CGDI) often start with the implementation of services related to metadata (discovery, evaluation, access).

However, the gathering of metadata is often a tedious and frustrating exercise that has to be performed on top of data management in spatial warehouses. Metadata standards are complex and cannot be readily applied to data managed in spatial warehouses where a data set can be defined as any object.

Intelec Geomatics has developed, under a GeoInnovations project a multistandard, multilingual metadata cataloguing (M^{3}Cat) tool (see www.intelec.ca/technologies). CubeWerx and SoftMap provide warehouse products to manage geospatial data. The three companies had an interest in improving metadata and data management in their products to facilitate their clients presence in geospatial data infrastructures.

This paper presents the results of the Integrated Web Catalogue Service (IWCatS) project realised under the GeoInnovations program.

2. CONTEXT

The project objectives were to develop tools that allow the integration of metadata servers in a Web environment using protocols similar to OpenGIS standards.

Figure 1 presents the context for an Integrated Web Catalogue service (IWCat). The context is based on the postulate that in geomatics, metadata are meant to help users discover and assess geospatial information and to help providers publish their data or services.

The model is composed of two objects and three actors[1]. The objects are metadata catalogues accessible through a Catalogue Server service and Spatial Warehouses accessible through a Feature Server service, while the actors are providers, librarians and consumers who have access to diverse services over the Web using standard protocols.

The consumer actor supports graphics and text based interactions by humans and automated agents. Consumers are users of both information found in metadata catalogues and spatial warehouses. Consumers browse and query catalogues in order to access or reference information of interest in spatial warehouses and browse metadata catalogues for more detail metadata information.

The metadata catalogue object includes information (who, what, where, etc.) and search engines to let users identify data of

---

[1] In some organisations the Librarian and the Provider may be the same person.

interest. Catalogues describe and reference content found in spatial warehouses, other catalogues or other data collections.

The spatial warehouse object contain data holdings comprised of feature collections, features and feature metadata. Spatial warehouses are offered by data providers.

The librarian actor manages catalogues. Librarians ingest, organize and extract information from spatial warehouses and create and augment catalogues.

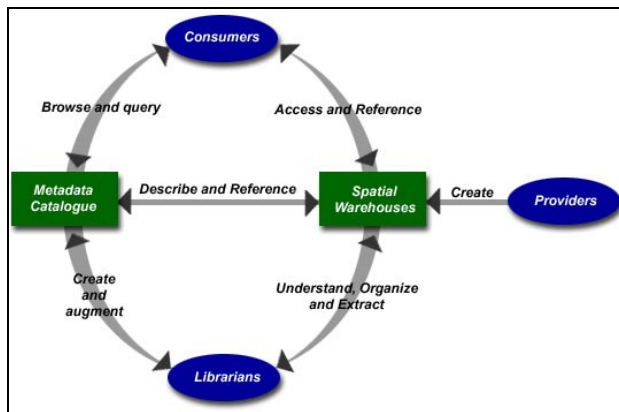The provider actor creates, updates and publishes spatial warehouses.

Figure 1. Context for the IWCat service

3. ISSUES AND ORIENTATIONS

To reach the project objectives, the project team had to resolve a number of issues, in particular:

- The level of integration between the metadata and spatial warehouses tools;
- The definition of granularity levels (data set versus object);
- The identification of common identifiers;
- The aggregation of metadata;
- The management of historical aspects;
- The development of protocols.

3.1 Level of Integration

There are two ways to integrate the management of metadata found in catalogues and spatial warehouses:

- to enable each service to interface with each other using open protocols;
- to integrate the management of metadata for discovery services in data warehouses tools.

Both approaches translate into a workable and operational environment. In theory, the second alternative may be preferable since the only reason that the two tools are not integrated is that for historical and performance reasons they have been developed as specific software.

However, the integration of the two (2) tools will not allow discovery services to work with data set collections not in spatial warehouses, e.g. files, paper products. And the integration of both tools will not correspond to the real world where each type of tool exists.

Hence, it was recommended to integrate the management of catalogue and spatial warehouse data by considering the two services as separate, but with the capability to access one service from another in a B2B fashion. Each service has the capability to be invoked by users of the other service. This will lead to global interoperability between the two services beyond proprietary services available from specific vendors.

This recommendation raised issues of security and availability (what happens when one of the services is down, how do we handle access privileges in both services) but these issues are common to any services.

3.2 Granularity

Metadata can describe geospatial data at different levels such as, geometric primitives, features, layers, maps, collections, data base, etc.

Some of these metadata fields are specific to each data set level while others are produced by performing operations on the lower levels metadata (geographic extent, content, dates, etc.).

In order to maintain an integrated environment for Web Catalogue Server (WCatS) services and Web Feature Server (WFS) services each service must define their respective metadata using a common encoding and format based on open standards such as FGDC or ISO 19115. Both the catalogue and spatial warehouse services must maintain a metadata model where the schema and the relations can be discovered over the Web using simple Web Catalogue Server interfaces. However with current metadata implementations (catalogue servers and spatial warehouses) the levels of granularity are often not identified uniquely in both services and the metadata elements they commonly manage do not have the same semantics.

Hence it was recommended that additional functionality be provided from an Integrated Web Catalogue Server environment perspective for maintaining and harvesting a metadata catalogue or data holding based on mapping different metadata models. The data holding is user configurable and contains for each level of granularity the metadata mapping between the catalogue and spatial warehouse environments :

- **The correspondence between the levels of granularity (e.g. level data set in catalogue correspond to level layer in spatial warehouse or feature level in spatial warehouse correspond to sub-data set in catalogue) ;**
- **The correspondence between metadata elements for each metadata standard maintained by the catalogue and spatial warehouse environments (e.g. ArchiveFormats in spatial warehouse corresponds to 6.4.2.2.2.3 Recording Format in a FGDC catalogue);**
- **How do these metadata elements react to one another (direct correlation, format conversion, etc.).**

The detailed structure of this data holding is presented in the data model (see Section 4).

3.3 Common Identifiers

For metadata management to be synchronized between the two environments, a method must be developed to uniquely identify and correlate, at each level of granularity, geospatial data objects for which metadata management must be integrated.

Ideally, the Spatial Warehouse unique identifier should be used in the Catalogue and the relations between the levels of granularity (primitive within a feature, feature within a layer, etc.) maintained by the warehouse.

However, relations between objects in each service are not always one to one:

- A catalogue can describe data in more than one spatial warehouse;
- Multiple catalogue entries could describe one spatial warehouse object (a feature or a pixel with different sources).

It was recommended that warehouse identifiers be used in the Catalogue. To take into account exceptions, a N to N table was also made available in the IWCat environment so that relations between each data occurrence is maintained when required. In all cases the relations between the geospatial data levels of granularity are maintained in the Spatial Warehouse.

3.4 Metadata Aggregation

Two different issues were considered:

- aggregation of metadata
- and metadata about aggregation.

An example of the first case is when one has the same metadata that are used at different levels of granularity, the metadata of the coarser level being derived from the metadata at the finer level (ex. date of last update for a data set being derived from the dates of the updates of object occurrences). When this is of interest to the user, one must check the integrity between the different levels of granularity.

It was recommended that procedures to do this automatically be implemented into the tools involved in the project as well as the corresponding updating rules.

An example of the second case is when one is interested to describe how global data have been derived from detailed data (ex. in multidimensional datacubes of warehouses, the total area of industrial land uses of a county is derived from the sum of the areas of the land uses of individual properties), presently no standard exist in geomatics to describe the aggregation algorithms and **this was not included in the scope of the project. However, it was recommended that the correspondence between metadata in spatial warehouses and in catalogues be considered a one to many relationship.**

3.5 Historical Aspects

We had to consider, if desired, a way to keep a trace of the evolution of the metadata (ex. for legal reasons). This can be done using the different versioning approaches familiar to spatio-temporal database design, however one must keep the synchronization between the different states of a data value and its metadata in the warehouse or the catalogue.

Services to define the requirements for the updating were not developed in the project (ex. frequencies or thresholds in each direction, format of the update metadata, synchronization rules).

3.6 Protocol

It is critical that geospatial data and metadata be maintained current and in synchronization and that the protocol used be flexible enough to support not only the access to metadata but perform all required transactions for the maintenance of metadata between the Catalogue and Spatial Warehouse services.

The on-going draft OGC Web Feature Server specification is both an encoding/ transport mechanism and a transaction mechanism for simple geographic features in XML/GML. The WFS interfaces supports INSERT, UPDATE, DELETE, QUERY and DISCOVERY operations on OGC Simple Feature datastores using HTTP as the distributed computing platform. Although the focus of the OGC WFS specification was a geometric feature and its attribution information there are no restrictions for the applicability of this protocol, or similar protocol, for metadata.

It was therefore recommended that the OGC Web Feature Server (WFS) interface specification be adapted for its applicability for the management of metadata between the Spatial Warehouse and the Catalogue services.

4. DATA

The data model for the Integrated Web Catalogue (IWCat) data store is presented on Figure 2. This data store has to be maintained in both the catalogue and spatial warehouse environments.

The project used XML for encoding metadata information and GML Profile 2 for the encoding of geographic information in XML.

5. OUTCOMES

The project outcomes will generate benefits for suppliers and users within spatial data infrastructures. Suppliers will have the capability to:

- Update their metadata catalogues from their data warehouses.
- Propagate metadata updates from a data set to all the objects it contains.
- Improve the connection of their metadata to the CGDI.

The project was a good example of partnership between the Federal Government, Provincial Governments, the Private Sector and Universities.
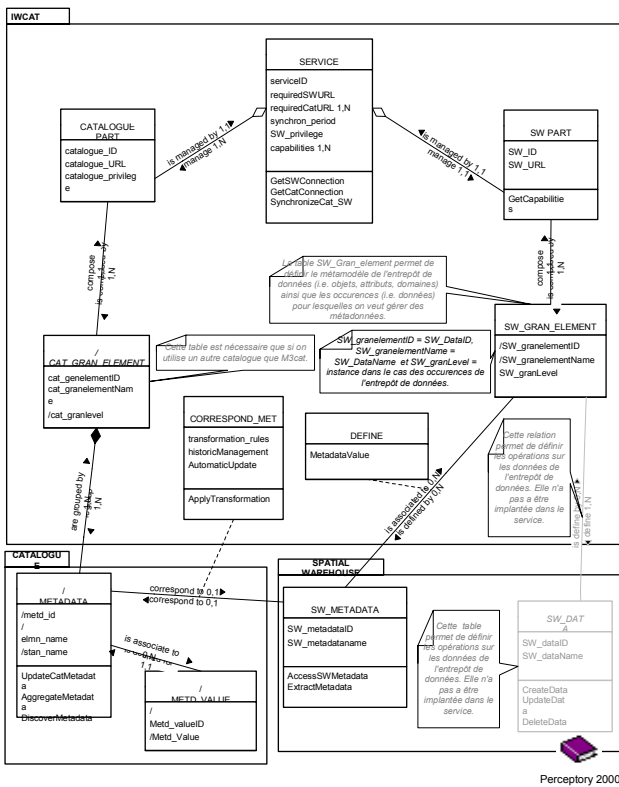

Figure 2: Data Model

Users will have the capability to:

- Query additional information on the data from data warehouses.
- Identify, when querying spatial data infrastructures, available data sets as well as specific data elements (concept of evolution from a data set to the information).

The project results will essentially allow suppliers to devote more time to the improvement of their data thanks to the economies of time obtained in the management of their metadata. These improvements could take the form of gathering of new data, better precision or more frequent updates.

The users are then able to better discover the data of interest, those will be of better quality and they will be better documented, which will ensure a more efficient use in their specific fields (forestry, mines, etc.).

6. CONCLUSION

The IWCat service was developed by Intelec Geomatics (Intelec) and its partners CuberWerx, SoftMap Technologies and the Centre for Research in Geomatics (CRG) of Laval University. It has been tested with six governmental organisations involved in the management of geospatial data: the Department of National Defence, Environment Canada, the Canadian Centre for Remote Sensing, the Canadian Forest Service, the Earth Sciences Sector of Natural Resources Canada and the Quebec Ministry of Natural Resources.